Technical Disclosure Commons

Defensive Publications Series

September 2020

Unsupervised Classification of Deconvolved Mass Spectra

Anonymous

Unsupervised Classification of Deconvolved Mass Spectra

ABSTRACT

This disclosure describes using self-organizing maps (SOMs) to perform unsupervised classification of deconvolved mass spectra for mass spectrometry applications.

KEYWORDS

- Self-organizing map
- Mass spectra
- Unsupervised classification
- Unsupervised learning
- Mass Spectrometry
- Mass Spectrometer
- Deconvolved
- SOM

BACKGROUND

A mass spectrometer is an analytical instrument used to measure the mass-to-charge ratios (m/z) of ions of a sample-under-analysis, or an analyte. Typically, the analyte is separated into components via a chromatographic instrument (e.g., via liquid chromatography, gas chromatography, or capillary electrophoresis), the separated components are introduced into an ion source of the mass spectrometer for ionization, and the resulting ions are subject to transport, confinement, and separation by the components of the mass spectrometer for analysis. The analysis can include generating a mass spectrum depicting a plot of intensity (relative abundance) as a function of the m/z. The mass spectrum is useful for the identification, quantification, and structural elucidation of the sample, for example, peptides, proteins, and related molecules.

Classification of deconvolved mass spectra is useful for understanding mass spectrometry data. In some scenarios, classification is performed “by hand,” involving examining peaks of the mass spectra and comparing the peaks with expected results, a model, or using supervised training techinques. In supervised training techniques, a set of labelled training data in which the results are previously known is used train a system to perform classifications. These scenarios require extensive human intervention, an appropriate set of models, and/or a set of training data.

DESCRIPTION

In contrast to supervised classification (or learning), unsupervised classification schemes perform classifications by measuring differences between data points with reference to any models or labelled training data. A self-organizing map (SOM) is an artificial neural network (ANN) trained using

unsupervised classification to identify undetected patterns in a data set with no pre-existing labels or low (or no) human direction. The SOM typically produces a two-dimensional representation of an input space using training samples, or a map, to transform the data set from a high-dimensional space into a low-dimensional space for better understanding of the data set.

As described herein, SOMs can be used to perform unsupervised classification of deconvolved mass spectra. The SOMs are used to classify unknown data samples based on identified similarities between their attributes without the need of a previously-classification training set. For example, the SOMs can map data samples from an N-dimensional sample space with each attribute represented by a corresponding dimension, into a lower-dimensional classification space in a manner that preserves topology, such that similar samples are mapped to similar regions in a classification space. This is depicted in Figure 1 below.

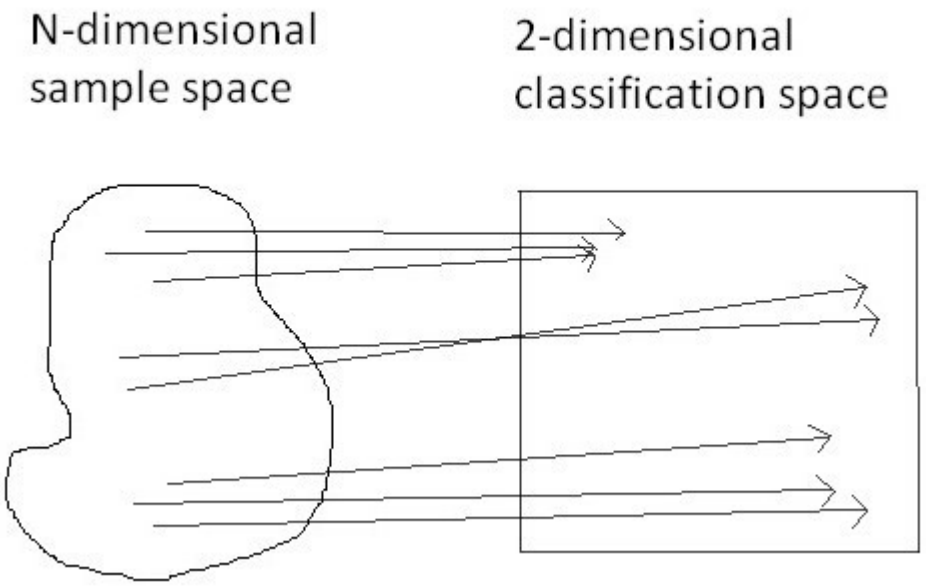

Figure 1

In more detail, a list of the most abundant masses (e.g., m/z) in each deconvolved mass spectrum is obtained, and a sliding window is applied to determine masses that are identical within an appropriate mass tolerance and identify a list of possible masses, as shown in Figure 2. This list of possible masses defines an N-dimensional parameter space.

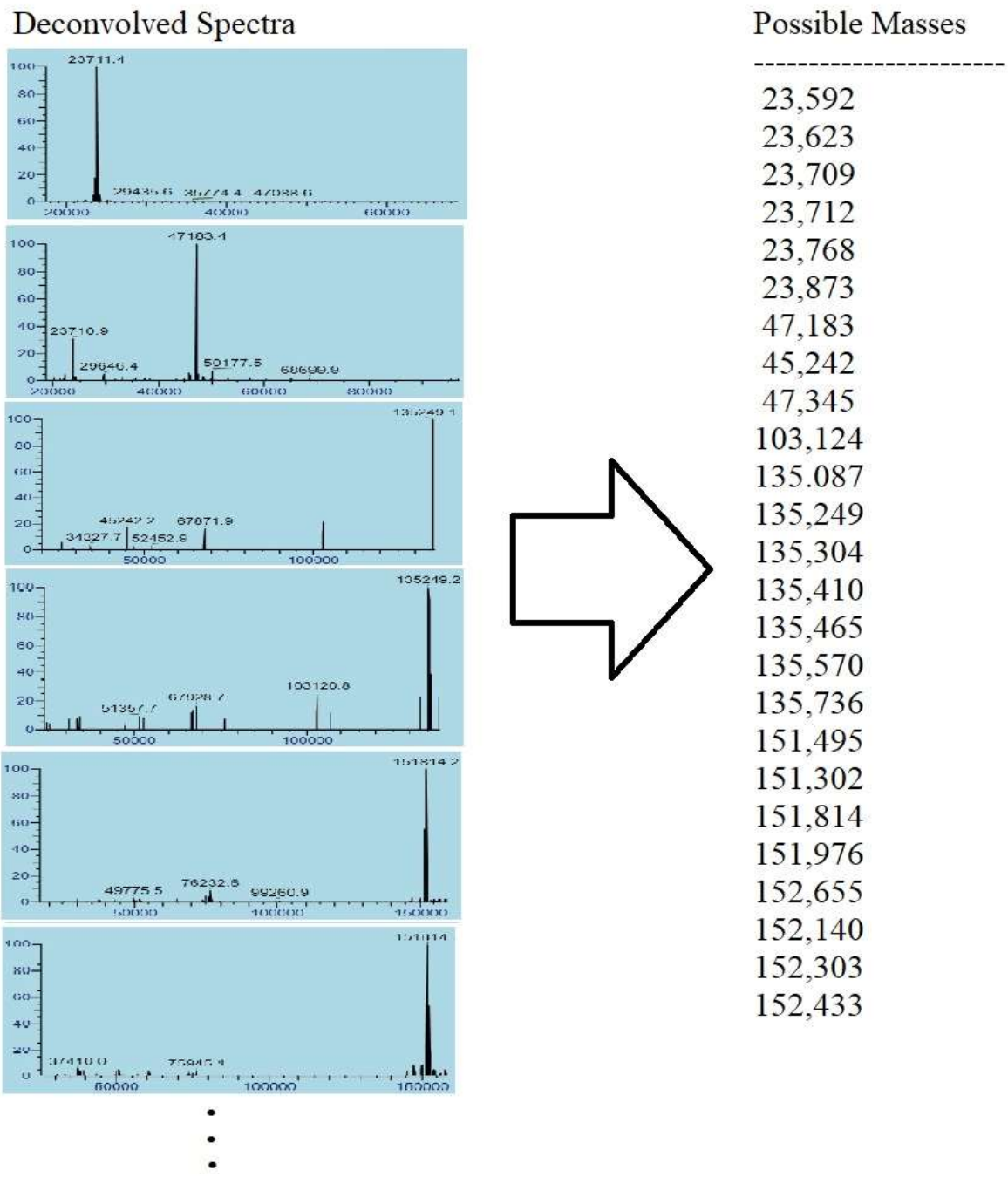

Figure 2

Each deconvolved spectrum is then converted to a list of abundances for the possible masses identified as describe above, and as shown in Figure 3. This represents each spectrum as a vector in the N-dimensional sample space.

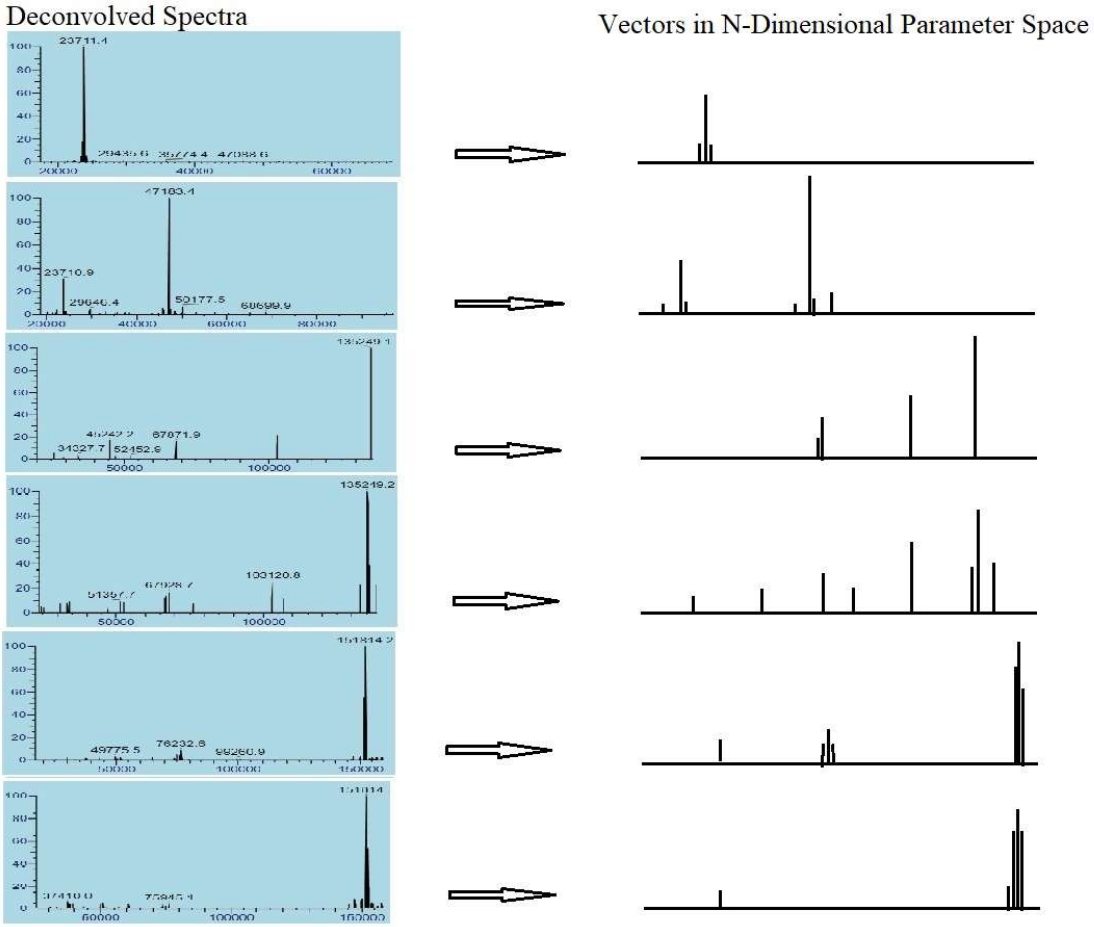

Figure 3

The vectors are then used to train a self-organizing map, as shown in Figure 4.

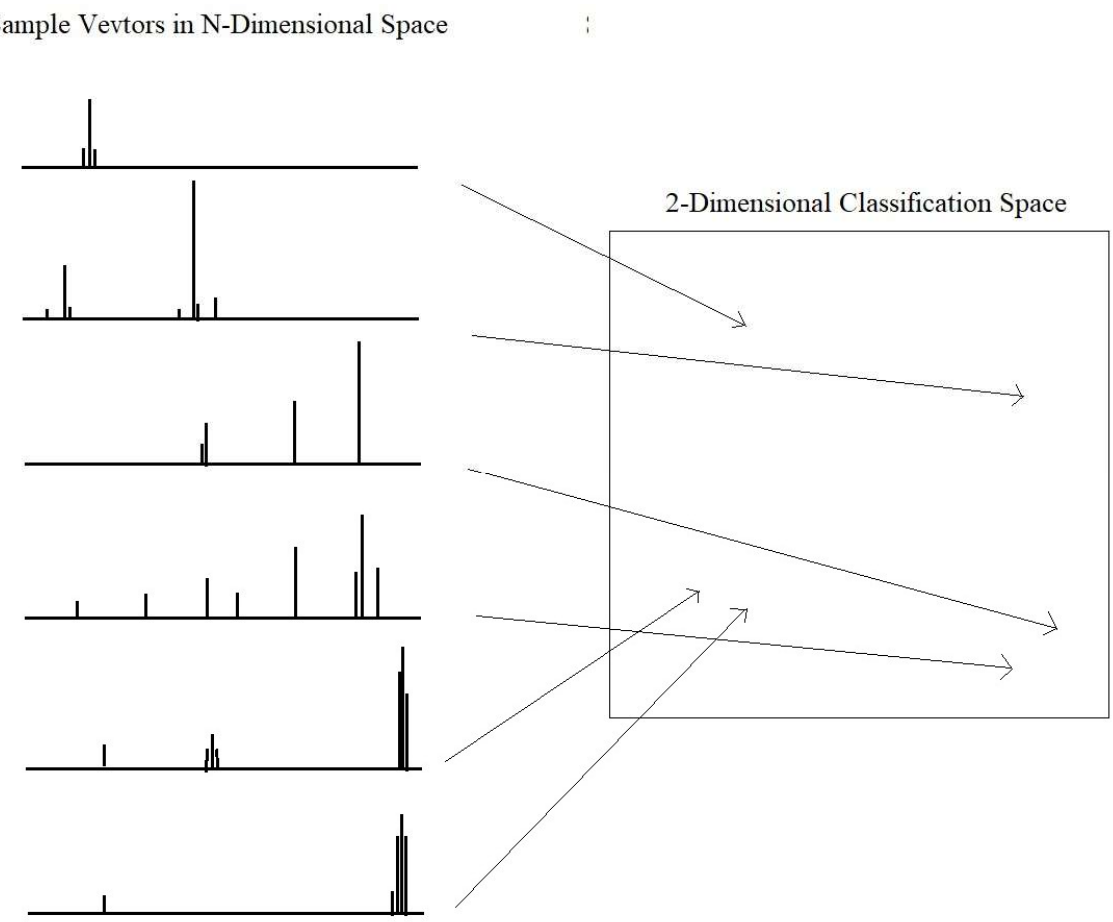

Figure 4

Use a partitioning scheme (e.g., u-vectors, minimal spanning trees, or Bayesian Blocks), regions of the SOM are identified and associated with possible classes, as shown in Figure 5 below.

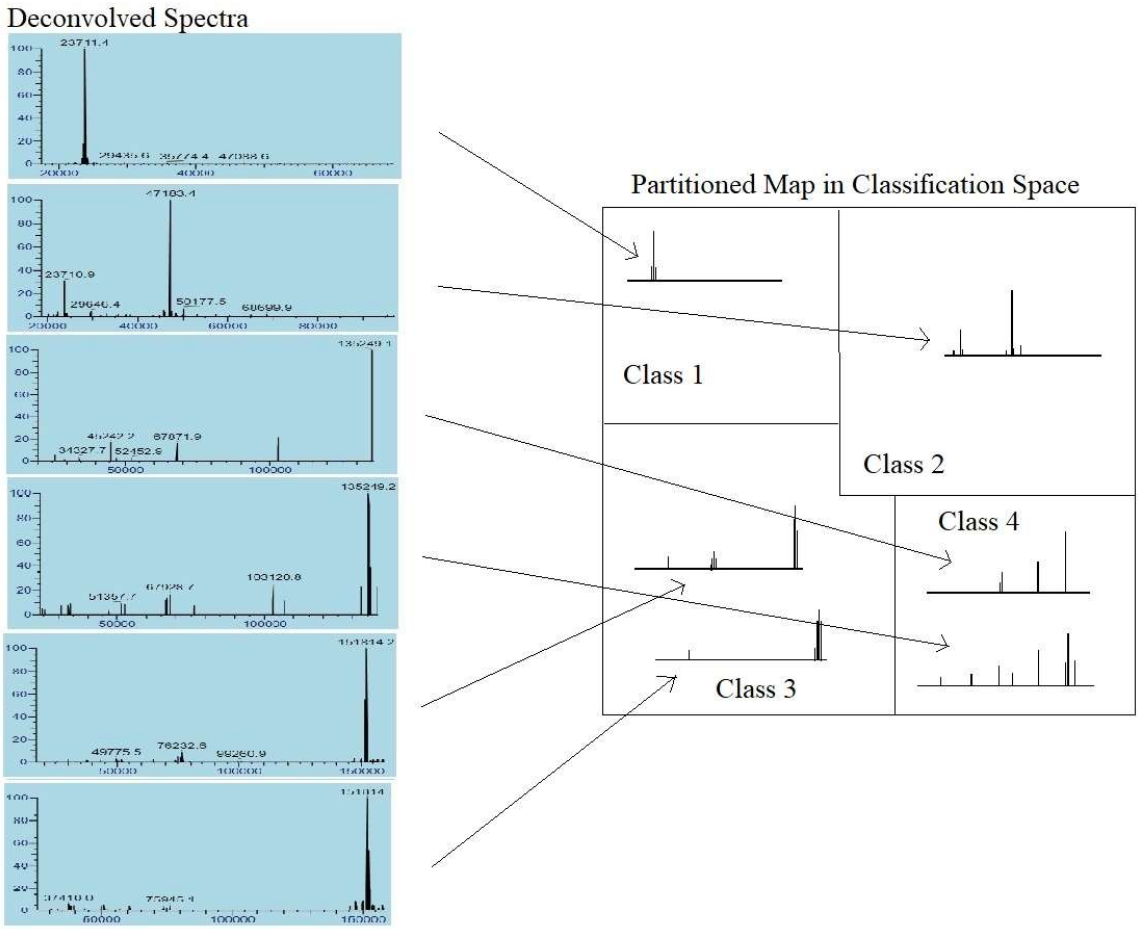

Figure 5

The partitioning scheme above may be adjusted to under-classify (e.g., group similar classes together) or over over-classify (e.g., split some classes) depending on the needs of the application.

After the self-organizing map has been trained and partitioned, as described above, it can be used to classify new deconvolved spectra by converting them to vectors in the N-dimensional sample space, as described above, and inputting these vectors to the trained map to determine what partition they're assigned to.

The map can also be 'seeded' with data for which the results are known to characterize the different partitions. Additionally, the distance between the assignments of two spectra to points in the map can be used as a measure of mass spectral similarity.

These techniques can be performed on a chromatography or mass spectrometry system, or using a computing system for post-acquisition analysis.

CONCLUSION

This disclosure describes the use of self-organizing maps (SOMs) to perform unsupervised classification of deconvolved mass spectra for mass spectrometry applications. Per the techniques of the disclosure, mass spectra can be classified to provide insight for an experiment.

REFERENCES

1. Kohonen, T. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69 (1982).